
















ORIGINAL

Deep learning-based computerized diagnosis of lung cancer

Diagnóstico computerizado del cáncer de pulmón basado en el aprendizaje profundo

Rakesh Sankaran¹  , Sheuli Sen²  , Lakshay Jeet Singh³  , Jaspreet Sidhu⁴  , Anisha Chaudhary⁵ 
  Jagtej Singh⁶  

¹Tagore Medical College and Hospital, Department of Radiodiagnosis. Chengelpet, India.

²Teerthakar Mahaveer University, Department of Nursing. Uttar Pradesh, India.

³GMCH, Department of MBBS. Chandigarh, India.

⁴Chitkara University, Centre of Research Impact and Outcome. Punjab, India.

⁵Quantum University. Uttarakhand, India.

⁶Chitkara University, Centre for Research and Development. Himachal Pradesh, India.

Cite as: Sankaran R, Sen S, Jeet Singh L, Sidhu J, Chaudhary A, Singh J. Deep learning-based computerized diagnosis of lung cancer. Salud, Ciencia y Tecnología. 2024; 4:.922. <https://doi.org/10.56294/saludcyt2024.922>

Submitted: 26-12-2023

Revised: 17-04-2024

Accepted: 21-08-2024

Published: 22-08-2024

Editor: Dr. William Castillo-González 

ABSTRACT

The Deep-Learning (DL) technique is capturing increasingly flexible in the sector of processing medical images. Rapid and precise lung cancer detection requires a standardized computer-aided diagnostic (CAD) architecture. For a quick and reliable detection of lung cancer, a standardized CAD framework is required. High-risk patients are advised by the National Lung Screening Trial to undertake standard screenings with low-dose CT to support the early detection of cancer and decrease the consequence of lung cancer death. In this paper, a lung CT scan and probabilistic bilateral convolutional neural networks (PB-CNN)-based automated diagnosis system for lung cancer are developed. The PB-CNN models were trained using sample cases from the LUNA16 dataset. We used existing techniques, such as Decision Trees (DT), Artificial Neural Networks (ANN) and K-Nearest Neighbors (KNN) to detect lung cancer. We employed accuracy, precision, recall, and f-measure in our experimental investigation. The proposed PB-CNN is automatically detecting lung cancer, yielding an acceptable performance.

Keywords: Lung Cancer Diagnosis; Computer Aided Diagnostics (CAD); Probabilistic Bilateral Convolutional Neural Networks (PB-CNN); Image Processing.

RESUMEN

La técnica Deep-Learning (DL) está adquiriendo cada vez más flexibilidad en el sector del procesamiento de imágenes médicas. La detección rápida y precisa del cáncer de pulmón requiere una arquitectura estandarizada de diagnóstico asistido por ordenador (CAD). Para una detección rápida y fiable del cáncer de pulmón se requiere una arquitectura CAD estandarizada. La National Lung Screening Trial recomienda a los pacientes de alto riesgo someterse a pruebas de detección estándar con TC de baja dosis para favorecer la detección precoz del cáncer y reducir las consecuencias de la muerte por cáncer de pulmón. En este artículo se desarrolla un sistema de diagnóstico automático del cáncer de pulmón basado en la TC pulmonar y en redes neuronales convolucionales bilaterales probabilísticas (PB-CNN). Los modelos PB-CNN se entrenaron utilizando casos de muestra del conjunto de datos LUNA16. Se utilizaron técnicas existentes, como árboles de decisión (DT), redes neuronales artificiales (ANN) y K-Nearest Neighbors (KNN) para detectar el cáncer de pulmón. En nuestra investigación experimental empleamos la exactitud, la precisión, la recuperación y la medida f. La PB-CNN propuesta detecta automáticamente el cáncer de pulmón con un rendimiento aceptable.

Palabras clave: Diagnóstico de Cáncer de Pulmón; Diagnóstico Asistido por Ordenador (CAD); Redes Neuronales Convolucionales Bilaterales Probabilísticas (PB-CNN); Procesamiento de Imágenes.

INTRODUCTION

The stage of cancer dictates a distance it spreads, and certain traits are required to distinguish malignant nodules from benign ones. These characteristics and combinations are used to evaluate cancer risk, although it can be difficult to link nodules to a definite diagnosis. Commonly used Computer-Assisted Diagnosis (CAD) techniques are based on features of cancer suspicion that have already been studied.⁽¹⁾ Cancer is seen as a serious problem that, due to ambiguous clinical tests and non-invasive therapies, may cause significant mortality in both women and men.⁽²⁾ The overall survival percentage of individuals with lung cancer is successfully increased by 20 % with early detection and appropriate therapy.^(3,4) A whole lung CT scan session typically consists of 150 to 300 images. Radiologists now do difficult diagnostic procedures, which test both their physical and mental stamina.⁽⁵⁾ To create a three-dimensional representation of an object's inside; digital geometry processing is used to combine several two-dimensional X-ray images obtained around a single axis of rotation. A lung X-ray is one of the most important diagnostic procedures for detecting lung cancer.⁽⁶⁾ Consequently, a technique to detect malignant nodules in their early stages is becoming more and more necessary.⁽⁷⁾

The optimal lung cancer treatment for each patient must also be chosen, which necessitates a precise diagnosis.⁽⁸⁾ In light of this, proposed PB-CNN for lung cancer diagnosis. Mohammed et al.⁽⁹⁾ merged CT images from the American Association of Physicists in Medicine and the Society of Photographic Instrumentations Engineers (SPIE-AAPM-LungX) collection to categorize lung nodules. Ozdemir et al.⁽¹⁰⁾ introduced an innovative computer-aided detection and diagnosis method that gives precise probability estimates for low-dose CT scan lung cancer screening. Poirier et al.⁽¹¹⁾ promoted the early detection of lung cancer. Out of twelve lung CT images, six were selected for healthy lungs and the other six were for those having scans from the Space Systems and Controls Lab (SCLC). Deshpande et al.⁽¹²⁾ contrasted the efficiency of applying deep organized algorithms to extract automatically generated features with the capacity to diagnose lung nodule CT images integrating traditional computer-aided diagnosis (CADx) methods involving features created by hand. Cong et al.⁽¹³⁾ focused attention on the benefits and drawbacks of DL in the pathophysiology of lung cancer and offered suggestions for the future. Joshi et al.⁽¹⁴⁾ analyzed the effectiveness of a DL-based method to quickly and correctly diagnose lung illness from lung cancer CT images. By utilizing widely used benchmark images, the Multilayer-CNN (M-CNN) was developed and proven to be effective at differentiating among different stages of lung cancer. Kasinathan et al.⁽¹⁵⁾ created the detection of tumors in the lungs utilizing Cloud-based Lung Tumor Detector and Stage Classifiers or Cloud-LTDSC was an active contouring modelling. Two methods were proposed by Meldo et al.⁽¹⁶⁾ to characterize the decisions generated by a system that employs computers to diagnose for lung cancer detection. Asuntha et al.⁽¹⁷⁾ utilized the input lung image, it was possible to classify lung cancer and its severity and find malignant lung nodules. Singh et al.⁽¹⁸⁾ established an effective method for identifying lung cancer from CT scan images and classifying them into benign and malignant categories. Meldo et al.⁽¹⁹⁾ examined a number of computer-aided techniques, assessed the most effective one presently in use, identified its shortcomings, and then suggested an updated version that enhanced. Manju et al.⁽²⁰⁾ employed image convolution, pooling, and Principal Component Analysis (PCA) methods to extract special patch-based characteristics. Wang et al.⁽²¹⁾ developed a CADx classification scheme to separate benign from malignant lung nodules. Rey et al.⁽²²⁾ developed a new technique for characterizing lung nodules on CT images and merging them into a single system to boost automation. The approach uses a difference-image method to suppress most of the normal features in the context and accentuate lung nodules.⁽²³⁾ Agnes et al.⁽²⁴⁾ developed to make it easier to filter expired air sample compositions by classifying all components into chemical groups and putting components into categories based on their occurrence levels of more than 75 % and higher than 90 %.⁽²⁵⁾

METHOD

The approach for diagnosing lung cancer in the suggested method involves a probabilistic bilateral convolutional neural network (PBCNN). Data collection is the initial stage in acquiring an input image for a successful diagnosis; database is accessible to public (<https://www.kaggle.com/datasets/avc0706/luna16>). A radiation beam is aimed toward the body using CT scanners, creating more accurate CT scan images. The CT volume data is generated utilizing the department of Digital Imaging and Communications in Medicine (DICOM) directory is then organized along with sequentially numbered. The widely used DICOM format for medical images makes it simpler for clinicians to access and diagnose patients. Because they give a detailed image for proper diagnosis, radiologists often study the two-dimensional (2D) and three-dimensional (3D) CT data in the axial, sagittal, and coronal planes. It makes sense to have a 2D perspective of people to better comprehend them since 3D CT scans of humans are complicated and include several anatomical features. The doctors can now reliably detect lung nodules on CT scans rather than lung X-rays to LIDC (lung image-database consortium) and image database-resource initiative or IDRI. Thoracic CT images with the indicated lesions are provided by the LIDC-IDRI collection, enabling research on lung cancer diagnosis utilizing CT imaging. Building the LIDC, an image repository, in collaboration with five educational foundations will allow international research into the creation of a CAD system to recognize lung lesions shown on CT images. Layered with axial lung cavity slice

sequences, the DICOM data from about a thousand patient records make up the LIDC-IDRI collection. The three types of detectable lesions are non-nodules, nodules less than 3 mm, and nodules larger than 3 mm. This set of information contains labeled and nodule-coordinated thoracic CT scan images and EML files.

Classification using Probabilistic Bilateral-Convolutional Neural Network (PB-CNN)

For image processing applications such as segmentation, denoising, and image filtering, a particular variety of CNN is the PB-CNN. The bidirectional filter used by PB-CNN as opposed to the conventional convolutional filter is an expansion of the conventional CNN design. The spatial separation and intensity differences between pixels are taken into consideration while calculating the filter response for the bilateral filter, which is a non-linear filter. Our motivation came from the current fascination with bilinear operation. Before a diagnosis is made, bilateral CNN reviews lateral and anterior-posterior (AP) X-ray images. Figure 1 conceptualizes PB-CNN.

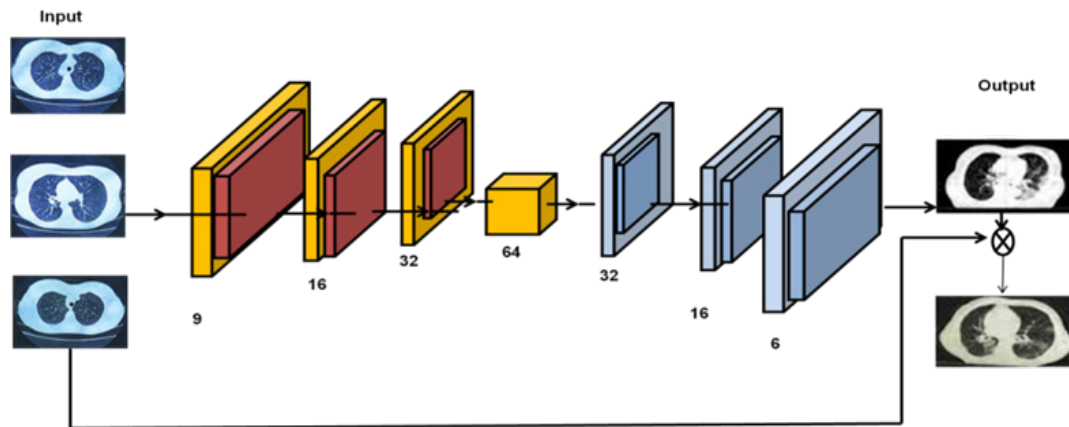


Figure 1. Schematic representation of PB-CNN

The patient x_i^c with i^{th} AP x-ray image was utilized to effectively create a bilinear classification. We purposely restructured the outputs of two PB-CNNs into $f_{(A)}(x_i) \mathbb{R}^{(D \times L)}$ and $f_{(B)}(x_i) \mathbb{R}^{(D \times L)}$, where D stands for the output dimensionality, L for their placements in space, and $f_{(.)}$. Convolutional layers for Visual Geometry Group (VGG-16) were used. The bilinear feature at point l might then be generated in the manner shown below equation (1).

$$\text{bilinear}(l, x_i, f_A, f_B) = f_A(l, x_i) f_B(l, x_i)^T \quad (1)$$

By combining the bilinear classification over every one place in these images and afterward factorizing them as follows, it is possible to generate the global bilinear representation $b_i^{\text{BCNN}} \in \mathbb{R}^{(D \times D)}$ equation (2).

$$b_j^{\text{BCNN}} = \sum_{l=1}^L \text{bilinear}(l, x_i, f_A, f_B) \quad (2)$$

By dividing the weights of $f_A(.)$ and $f_B(.)$, we were still able to achieve equivalent performance in comparison to the cases where $f_A \neq f_B$ was independent. The bilinear representation $b_i^{\text{BCNN}} \in \mathbb{R}^{(D \times D)}$ was then established as equation (3).

$$b_i^{\text{single}} = \sum_{l=1}^L \text{bilinear}(l, x_i, f_A, f_A) \quad (3)$$

They intended the bilinear classification of every image in addition to element-wise integrating them to construct the Two-path bilinear classification, as indicated in the following formula (4), to integrate the data from two distinct x-ray images of patient i that were obtained from different angles.

$$b_i^{\text{two}} = \sum_{l=1}^L \text{bilinear}(l, x_i^c, f_A, f_B) + \sum_{l=1}^L \text{bilinear}(l, x_i^l, f_B, f_B) \quad (4)$$

To produce a bilateral categorization that matched the two x-ray images shared Regions of Interest (ROI), the i^{th} patient's lateral x-ray and AP images (x_i^c and x_i^l) were used. This bilateral bilinear representation $b_i^{\text{bilateral}} \in \mathbb{R}^{(D \times D)}$ might be obtained specifically as follows equation (5).

$$b_i^{bilateral} = \sum_{l=1}^L \text{bilateral}(l, x_i^c, x_i^l, f_A, f_B) = \sum_{l=1}^L f_A(l, x_i^c) f_B(l, x_i^l)^T \quad (5)$$

While highlighting the unique characteristics of the images required for NF1-S diagnosis, AP and lateral X-ray image features yielded features with outermost product significant values that indicated similar regions of concern.

RESULTS

These experimental outcomes show how the suggested approach, PB-CNN, compares to the existing methods, ANN, KNN, and DT ⁽²⁵⁾. Results were based on accuracy, precision, recall, and the F-measure parameter. The degree to which a test or measurement properly detects or eliminates a certain ailment or disease is referred to as accuracy. Accuracy in the context of lung cancer may relate to diagnostic tests or imaging techniques' capacity to accurately determine a patient's lung cancer status.

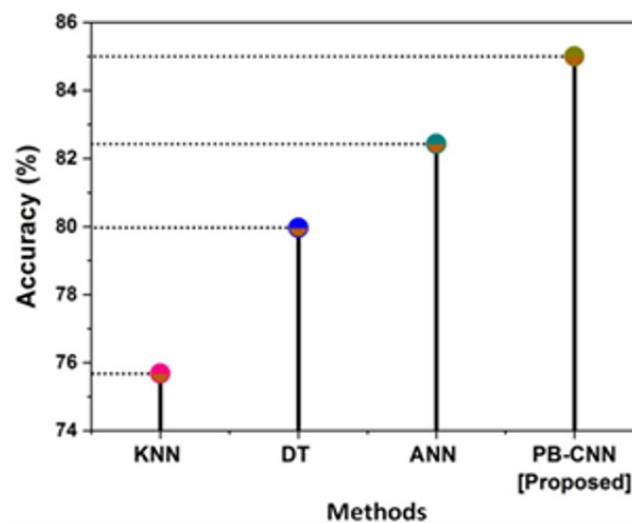


Figure 2. Comparative analysis of accuracy

Figure 2 compares the proposed PB-CNN approach to an existing method like KNN (75,68 %), DT (79,97 %), and ANN (82,43 %), and shows that PB-CNN (85,00 %) has higher accuracy. Precision in the context of lung cancer may refer to a diagnostic test or imaging technique's ability to reliably identify or rule out lung cancer in patients, even when carried out by various healthcare experts or using various tools.

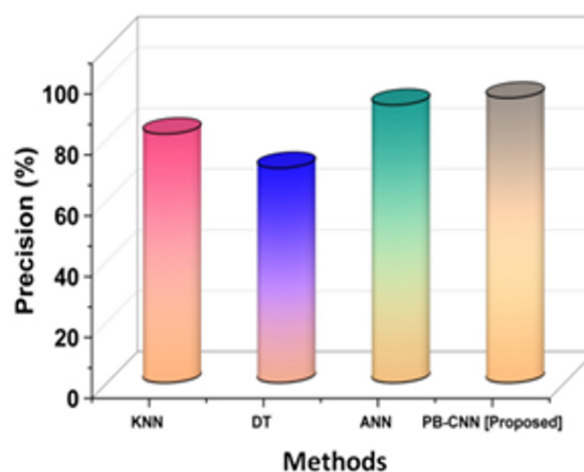


Figure 3. Analysis of precision

Figure 3 illustrates that PB-CNN (93,60 %) has superior precision when comparing the proposed PB-CNN methodology with existing methods like KNN (81,82 %), DT (70,59 %), and ANN (91,30 %). Recall, commonly referred to as sensitivity, is a statistical metric that expresses how well a diagnostic procedure or imaging method can identify individuals who have a certain illness or condition.

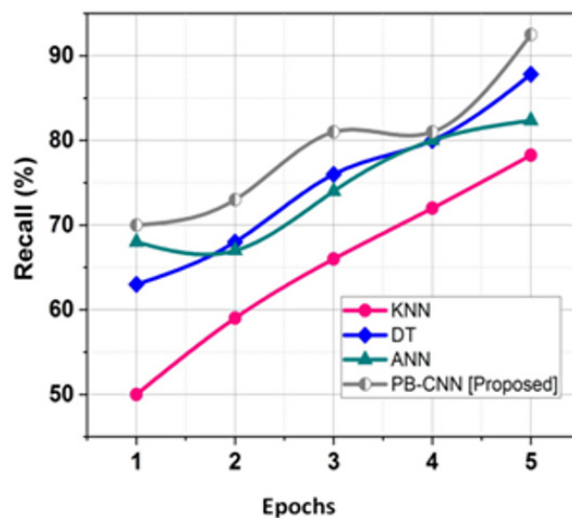


Figure 4. Comparative analysis of recall

Figure 4 compares the proposed PB-CNN technique to current approaches like KNN (78,26 %), DT (87,81 %), and ANN (82,35 %), demonstrating that PB-CNN (85,45 %) has a greater recall. A statistical metric called the f-measure combines precision and recall providing a comprehensive evaluation of how well a diagnostic test or classifier performed.

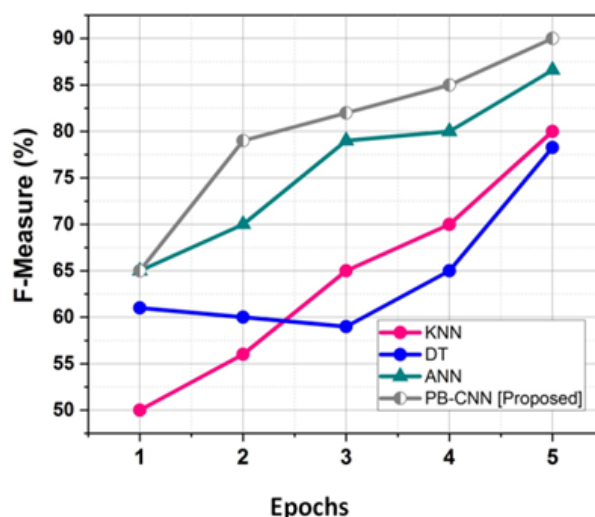


Figure 5. Analysis of F-measure

Figure 5 shows that the suggested PB-CNN strategy (90,00 %) has a higher F-measure when compared to popular methods like KNN (80,00 %), DT (78,26 %), and ANN (86,60 %).

DISCUSSION

KNN, ANN, and DT, along with other existing approaches, have all been discussed. The K-Nearest Neighbors (KNN) technique is a distance-based one; hence it takes a long time since it must compute the distance between each existing point and each new point. Data outliers may throw off KNN. The accuracy of categorization results may be significantly impacted by outliers, which is especially problematic in medical diagnostic applications. Computes at a high cost: When processing large datasets or a large number of features, KNN might be computationally expensive. One of the limitations of decision trees is that they are not as good at predicting the outcome of a continuous variable. Using decision trees to make predictions is difficult since even modifications to the data might produce a totally different tree structure. If decision trees are too weighted toward features with a high number of categories or those that are highly correlated with the dependent variable, the performance of the tree may suffer. One drawback of ANNs is that they can't be used until they've been trained. Over-fitting occurs when ANNs remember the training data and fail to effectively generalize to

fresh input. Although measures like early halting and regularization may help, this problem may remain. A large neural network is very time-consuming to process. PBCNN's benefits include quick and accurate image processing, among others. The problems of existing methods are addressed by PBCNN. That PBCNN may function more effectively.

CONCLUSIONS

The lung cancer is third in frequency overall. It is caused by lung cancer cells growing out of control. Treatment options include immunotherapy, radiation therapy, surgery, chemotherapy, and targeted medicines. Due to the late stages of the illness, over 75 % of those diagnosed with lung cancer do not qualify as prospects for surgery at their point of detection. PB-CNN based automated diagnosis for lung cancer is developed by lung CT images. Our testing results show that f-measure, accuracy, and precision outperform existing methods. PB-CNN is a complex model that requires significant computational resources and training time to achieve good performance. This can limit its overall usefulness in applications where the input data may vary widely. The focus of future research is on building PBCNNs that can withstand distributional changes, either by directly simulating the distribution or by using techniques like domain adaptation or meta-learning.

BIBLIOGRAPHIC REFERENCES

1. SU, Aswathy, et al. "Deep learning-based BoVW-CRNN model for lung tumor detection in nano-segmented CT images." *Electronics* 12.1 (2022): 14.
2. Pradhan, Kanchan Sitaram, Priyanka Chawla, and Rajeev Tiwari. "HRDEL: High ranking deep ensemble learning-based lung cancer diagnosis model." *Expert Systems with Applications* 213 (2023): 118956.
3. Huang, Shigao, et al. "Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective." *Seminars in Cancer Biology*. Vol. 89. Academic Press, 2023.
4. Kim, Jeffrey, Hobart Lee, and Brian W. Huang. "Lung cancer: diagnosis, treatment principles, and screening." *American family physician* 105.5 (2022): 487-494.
5. Lei, Juntian, et al. "An exercise prescription for patients with lung cancer improves the quality of life, depression, and anxiety." *Frontiers in Public Health* 10 (2022): 1050471.
6. Bhandary, Abhir, et al. "Deep-learning framework to detect lung abnormality-A study with chest X-Ray and lung CT scan images." *Pattern Recognition Letters* 129 (2020): 271-278.
7. Lakshmanaprabu, S. K., et al. "Optimal deep learning model for classification of lung cancer on CT images." *Future Generation Computer Systems* 92 (2019): 374-382.
8. Nooreldeen, Reem, and Horacio Bach. "Current and future development in lung cancer diagnosis." *International journal of molecular sciences* 22.16 (2021): 8661.
9. Mohammed, Shivan HM, and Ahmet Çinar. "Lung cancer classification with convolutional neural network architectures." *Qubahan Academic Journal* 1.1 (2021): 33-39.
10. Ozdemir, Onur, Rebecca L. Russell, and Andrew A. Berlin. "A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans." *IEEE transactions on medical imaging* 39.5 (2019): 1419-1429.
11. Poirier, John T., et al. "New approaches to SCLC therapy: from the laboratory to the clinic." *Journal of Thoracic Oncology* 15.4 (2020): 520-540.
12. Deshpande, Pallavi, et al. "Combining handcrafted features and deep learning for automatic classification of lung cancer on CT scans." *Journal of Artificial Intelligence and Technology* 4.2 (2024): 102-113. Meskina, Elena R. "Preliminary clinical and epidemiological analysis of the first 1,000 pediatric COVID-19 cases in Moscow Region." *Journal of microbiology, epidemiology and immunobiology* 97.3 (2020): 202-213.
13. Cong, Lei, et al. "Deep learning model as a new trend in computer-aided diagnosis of tumor pathology for lung cancer." *Journal of Cancer* 11.12 (2020): 3615.

14. Joshi, Shubham, et al. "Analysis of Smart Lung Tumour Detector and Stage Classifier Using Deep Learning Techniques with Internet of Things." *Computational Intelligence and Neuroscience* 2022.1 (2022): 4608145.
15. Kasinathan, Gopi, and Selvakumar Jayakumar. "Cloud-Based Lung Tumor Detection and Stage Classification Using Deep Learning Techniques." *Biomed research international* 2022.1 (2022): 4185835.
16. Meldo, Anna, et al. "The natural language explanation algorithms for the lung cancer computer-aided diagnosis system." *Artificial intelligence in medicine* 108 (2020): 101952.
17. Asuntha, A., and Andy Srinivasan. "Deep learning for lung Cancer detection and classification." *Multimedia Tools and Applications* 79.11 (2020): 7731-7762.
18. Singh, Gur Amrit Pal, and P. K. Gupta. "Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans." *Neural Computing and Applications* 31.10 (2019): 6863-6877.
19. Meldo, Anna, et al. "The natural language explanation algorithms for the lung cancer computer-aided diagnosis system." *Artificial intelligence in medicine* 108 (2020): 101952.
20. Manju, B. R., V. Athira, and Athul Rajendran. "Efficient multi-level lung cancer prediction model using support vector machine classifier." *IOP Conference Series: Materials Science and Engineering*. Vol. 1012. No. 1. IOP Publishing, 2021.
21. Wang, Bin, et al. "A fast and efficient CAD system for improving the performance of malignancy level classification on lung nodules." *IEEE Access* 8 (2020): 40151-40170.
22. Rey, Alberto, Bernardino Arcay, and Alfonso Castro. "A hybrid CAD system for lung nodule detection using CT studies based in soft computing." *Expert Systems with Applications* 168 (2021): 114259.
23. Zheng, Shaohua, et al. "Interpretative computer-aided lung cancer diagnosis: from radiology analysis to malignancy evaluation." *Computer Methods and Programs in Biomedicine* 210 (2021): 106363.
24. Agnes, S. Akila, and J. Anitha. "Appraisal of deep-learning techniques on computer-aided lung cancer diagnosis with computed tomography screening." *Journal of Medical Physics* 45.2 (2020): 98-106.
25. Günaydin, Özge, Melike Günay, and Öznur Şengel. "Comparison of lung cancer detection algorithms." 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). IEEE, 2019.

FINANCING

None.

CONFLICT OF INTEREST

None.

AUTHORSHIP CONTRIBUTION

Conceptualization: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Data curation: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Formal analysis: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Research: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Methodology: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Project management: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Resources: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Software: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Supervision: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Validation: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Display: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Drafting - original draft: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.

Writing - proofreading and editing: Rakesh Sankaran, Sheuli Sen, Lakshay Jeet Singh, Jaspreet Sidhu, Anisha Chaudhary, Jagtej Singh.